Claims

What is claimed is:

5.

1

2

| 1. | A method of assembling a life science knowledge base comprising the steps of: |
|----|---|
| | (a) generating two or more nodes indicative of life science data using |
| | a life science taxonomy; |
| | (b) assigning to one or more pairs of nodes a representation descriptor, |
| | the representation descriptor corresponding to a relationship between a pair of |
| | nodes; |
| | (c) assembling two or more nodes and one or more representation |
| | descriptors assigned to one or more pairs of said two or more nodes into an |
| | electronic database such that at least one of said two or more nodes is joined to a |
| | another node by a representation descriptor. |
| 2. | The method of claim 1 further comprising the step of receiving life science data, |
| | wherein the step of generating two or more nodes is based at least in part on said |
| | received life science data. |
| 3. | The method of claim 2, wherein the step of receiving life science data comprises |
| | collecting said life science data using a software agent. |
| 4. | The method of claim 2, wherein the step of receiving life science data further |
| | comprises receiving one or more of metadata and context data. |
| | 2. |

The method of claim 1, wherein said life science data comprises information

representative of a molecule, biological structure, physiological condition, trait,

- phenotype, biological process, clinical data, medical data, or disease data and
 chemistry.
- The method of claim 1, wherein said life science data comprises a descriptor of the condition, location, amount, or substructure of a molecule, biological structure, physiological condition, trait, phenotype, biological process, clinical data, medical data, or disease data and chemistry.
- 7. The method of claim 1, wherein the step of generating two or more nodes comprises reformatting at least a portion of said life science data.
- 1 8. The method of claim 1, wherein one or more of the representation descriptors 2 correspond to a epistemological relationship between a pair of nodes.
- 1 9. The method of claim 1, wherein one or more of the representation descriptors
 2 comprise a case frame.
- 1 10. The method of claim 1 further comprising the step of providing an ontology for
 2 use with representation descriptors, wherein the step of assigning to one or more
 3 pairs of nodes a representation descriptor is based on said ontology.
- 1 11. The method of claim 1, further comprising the step of segregating said electronic
 2 database into two or more sectors such that access may be restricted to one or
 3 more selected sectors.
- 1 12. The method of claim 1, wherein at least one of the two or more nodes itself
 2 represents a representation descriptor.

- 1 13. An article of manufacture having a computer-readable program carrier with
 2 computer-readable instructions embodied thereon for performing the method of
 3 claim 1.
- 1 14. A system for assembling a life science knowledge base comprising:

5

7

8

9

- 2 (a) a data collector configured to receive life science data and to 3 generate nodes based on said life science data;
 - (b) a relationship generator configured to assign a relationship descriptor to a pair of nodes, the representation descriptor corresponding to a relationship between a pair of nodes; and
 - (c) a knowledge assembler configured to assemble two or more nodes and one or more representation descriptors assigned to one or more pairs of said two or more nodes into an electronic database such that at least one of said two or more nodes is joined to another node by a representation descriptor.
- 1 15. The system of claim 14 further comprising a graphical user interface configured
 2 to permit a user to query the electronic database at least on the relationship
 3 between at least two nodes.
- 1 16. The system of claim 14 further comprising a data input interface configured to
 2 permit a user to submit life science data to the data collector.
- 1 17. The system of claim 16, wherein the data input interface is further configured to
 2 permit a user to assign a representation descriptor to a pair of nodes in the
 3 electronic database.

- 1 18. The system of claim 14 further comprising an access manager configured to 2 restrict access of a user to one or more portions of the electronic database.
- The system of claim 14 further comprising a software agent in electronic communication with the data collector, wherein the software agent is configured to collect life science data.
- The system of claim 14, wherein said life science data comprises information representative of a molecule, biological structure, physiological condition, trait, phenotype, biological process, clinical data, medical data, or disease data and chemistry.
- The system of claim 14, wherein said life science data comprises a descriptor of the condition, location, amount, or substructure of a molecule, biological structure, physiological condition, trait, phenotype, biological process, clinical data, medical data, or disease data and chemistry.
- The system of claim 14 further comprising a library of machine readable representation descriptors in electronic communication with the relationship generator.
- The system of claim 14, wherein one or more of the representation descriptors correspond to a epistemological relationship between a pair of nodes.
- The system of claim 14, wherein one or more of the representation descriptors comprise a case frame.

25. The system of claim 14, wherein one or more of the nodes represents a 1 2 representation descriptor. 26. A computer program product comprising: an electronic database storing a plurality of case statements, each case 2 statement comprising; 3 4 an first object identifier; a relationship connector; and 5 a second object identifier 6 wherein the relationship connector is based on a life science ontology. 7 27. The product of claim 26, wherein a set of said case statements define a biological function. 2 28. 1 The product of claim 27, wherein the biological function comprises a chemical reaction. 2 29. The product of claim 27, wherein the biological function comprises transport. 1 30. The product of claim 27, wherein the biological function comprises digestion of a 1 biomolecule. 2 31. The product of claim 26, wherein at least one of the first and second object 1 identifiers identifies a biomolecule. 2

The product of claim 26, wherein at least one of the first and second object

identifiers identifies a biological function.

32.

1

- 1 33. The product of claim 26, wherein at least one of the first and second object identifiers identifies a relationship connector.
- 1 34. The product of claim 26, wherein a the relationship connector represents an identity relationship.
- 1 35. The product of claim 26, wherein a the relationship connector represents a product relationship.
- 1 36. The product of claim 26, wherein a the relationship connector represents a substrate relationship.
- 1 37. The product of claim 26, wherein a the relationship connector represents a enzymatic relationship.
- The product of claim 26 further comprising a graphical user interface configured to permit a user to query the database at least on the relationship between biological object identifiers.
- 1 39. The product of claim 26 further comprising a data input interface configured to
 2 permit a user to create case statements.
- 1 40. The product of claim 26 further comprising an access manager configured to
 2 restrict access of a user to one or more portions of the electronic database.
- 1 41. A method of providing knowledge about and permitting investigation of a 2 functional physicochemical life science pathway comprising the steps of:

 (a) generating plural nodes indicative of life science data using a life science taxonomy;

- (b) assigning to respective pairs of nodes case frames corresponding to a physical or functional relationship between the node pairs;
- (c) assembling the nodes and case frames to produce an addressable electronic life science knowledge base representing a web of nodes interrelated by case frames; and
- (d) providing a user interface including a knowledge base search engine to permit a user to locate and obtain data descriptive of a selected region of said web.
- 1 42. A method of converting life science data into life science knowledge comprising
 2 the steps of:
 - (a) providing an electronically accessible knowledgebase representative of a web of life science information comprising a plurality of nodes indicative of life science data interrelated by a plurality of case frames representative of the functional or physical relationship between pairs of nodes; and
 - (b) providing through an electronic link to a plurality of life science data providers a data input interface which permits diverse providers to input new data into said knowledgebase, requires input of the new data to fit within one or more of a plurality of predetermined node categories, and requires input of the new data to fit within one or more of a plurality of case frames, thereby to enforce

a consistent data structure and ontology rules and to permit iterative accumulation 13 of new data linked functionally within said knowledgebase. 14 1 43. A method for representing life science knowledge, the method comprising using a 2 case frame to represent the life science knowledge, the case frame comprising: a first object identifier; 3 a relationship connector; and 4 a second object identifier, 5 wherein the relationship connector is based on a life science ontology. 6 44. The method of claim 43, wherein the case frame represents a biochemical reaction. 2 45. The method of claim 44, wherein the biochemical reaction is an enzymatic 1 2 reaction. 46. The method of claim 43, wherein the case frame represents binding. 1 47. The method of claim 46, wherein the binding comprises nucleic acid binding. 48. The method of claim 43, wherein the case frame represents an agglomeration of 1 molecules. 2 49. The method of claim 43, wherein the case frame represents a modification of a 1 2 polymer. The method of claim 49, wherein the modification of a polymer comprises a post-50.

translational modification of a protein.

- 1 51. The method of claim 50, wherein the post-translational modification is selected
- 2 from a group consisting of phosphorylation, acetylation, peptide-bond cleavage,
- glycosylation, lipidation, fatty-acylation, prenylation, methylation, metallation,
- 4 cross-linking, hydroxylation, sulfation, ADP-ribosylation, and covalent
- 5 attachment of prosthetic groups.
- 1 52. The method of claim 43, wherein the case frame represents competition between
- 2 processes.
- 1 53. The method of claim 43, wherein the case frame represents requirements of a
- 2 process.
- 1 54. The method of claim 43, wherein the case frame represents translocation.
- 1 55. The method of claim 43, wherein the case frame represents activation.
- 1 56. The method of claim 55, wherein the case frame represents direct activation.
- 1 57. The method of claim 43, wherein the case frame represents inhibition.
- 1 58. The method of claim 57, wherein the case frame represents direct inhibition.
- 1 59. The method of claim 43, wherein the case frame represents patient data.
- 1 60. The method of claim 59, wherein the case frame represents patient data of a
- 2 phynotypic nature.
- 1 61. The method of claim 59, wherein the case frame represents patient data of a
- 2 genotypic nature.

- 1 62. The method of claim 43, wherein the case frame represents a product relationship.
- 1 63. The method of claim 43, wherein the case frame represents an experiment.
- 1 64. The method of claim 63, wherein the relationship connector identifies linkages to 2 characteristics and parameters of the experiment.
- 1 65. The method of claim 43, wherein the case frame represents an animal model.
- 1 66. The method of claim 43, wherein the case frame represents a tissue type.
- 1 67. The method of claim 43, wherein the case frame represents a SNP.
- 1 68. The method of claim 43, wherein the case frame represents a splice varient.
- 1 69. The method of claim 43, wherein the case frame represents microRNA.
- 1 70. The method of claim 69, wherein the relationship connector identifies linkages to a function of the microRNA.
- The method of claim 43, wherein the first object identifier represents a chromosome.
- The method of claim 43, wherein the case frame represents a chromosomal modification.
- 1 73. The method of claim 43, wherein the case frame represents a silenced gene.
- The method of claim 43, wherein the case frame represents a subcellular compartment.

- 1 75. The method of claim 43, wherein the relationship connector represents a subset
- 2 relationship.
- 1 76. The method of claim 43, wherein the relationship connector represents a
- 2 positional relationship.
- 1 77. The method of claim 43, wherein the relationship connector identifies a drug
- 2 combination for treatment of one or more conditions.
- 1 78. The method of claim 43, wherein the case frame represents a disease.
- 1 79. The method of claim 78, wherein the disease is cancer.
- 1 80. The method of claim 43, wherein the case frame represents a population.
- 1 81. The method of claim 43, wherein the case frame represents a xenograft model.
- 1 82. The method of claim 43, wherein the case frame represents a cell line.
- 1 83. The method of claim 43, wherein the relationship connector represents that an
- 2 entity is a biomarker.
- 1 84. The method of claim 43, wherein the first object identifier identifies a protein.
- 1 85. The method of claim 43, wherein the first object identifier identifies a gene.
- 1 86. The method of claim 43, wherein the first object identifier identifies a metabolite.
- 1 87. The method of claim 43, wherein the first object identifier identifies a population.

- 2 88. The method of claim 43, wherein the first object identifier identifies an epidemic.
- 1 89. The method of claim 43, wherein the first object identifier identifies a pathogen.
- 1 90. The method of claim 89, wherein the pathogen is selected from a group consisting of a virus, a bacteria, a fungus, and a prion.
- 1 91. The method of claim 89, wherein the relationship connector represents the
- 2 implications of a specific disease.